ERDC TN-DOER-C23
September 2001

Review of Field Validation Studies of Sediment Bioassays for the Regulatory Evaluation of Dredged Material

PURPOSE: This technical note reviews published field validation studies of sediment bioassays to recommend the design and conduct of such studies in the future. The primary impetus for this review is to develop an approach for the validation of emerging chronic sublethal sediment toxicity tests for the regulatory evaluation of dredged material.

BACKGROUND: The U.S. Army Corps of Engineers (USACE) has statutory authority for evaluating the potential unacceptable, adverse impacts of dredged material disposal. Regulations implementing Section 103 of the Marine Protection, Research, and Sanctuaries Act (PL 92-532) state that "material shall be deemed environmentally acceptable for ocean dumping only when...no significant undesirable effects will occur due either to chronic toxicity or to bioaccumulation...." Similarly, regulations implementing Section 404(b) 1 of the Clean Water Act (PL 92-500) state that "the permitting authority shall determine in writing the potential short-term or long-term effects of a proposed discharge of dredged or fill material on the physical, chemical, and biological components of the aquatic environment...."

Beginning in the late 1970's and early 1980's USACE in cooperation with the U.S. Environmental Protection Agency (USEPA) developed an effects-based approach for the regulatory evaluation of dredged material. Ultimately, two guidance manuals were produced, the Ocean Testing Manual and the Inland Testing Manual (USEPA/USACE 1991 and 1998, respectively). The approach relied on the use of sediment bioassays to determine the suitability of the material for aquatic disposal. The bioassays implemented for the regulatory program included solid-phase (SP) bioassays to predict the potential toxicity of the dredged material on benthic invertebrates at the disposal site and suspended particulate phase (SPP) bioassays, also known as sediment elutriate tests, to evaluate potential water column toxicity shortly after disposal outside the mixing zone. Both SP and SPP bioassays focused on acute effects (e.g., typically 10 days for SP tests and 96 hours for SPP tests) with survival as the primary end point. In addition to these acute benthic and water column toxicity tests, tests to assess the bioaccumulation potential of sediment-associated contaminants were also included in the regulatory program (USEPA/USACE 1991, 1998).

While the regulations governing the statutory authority of the Corps mandated consideration of "chronic" and "long-term effects," the tools for evaluating those effects in sediment systems were not adequately developed for regulatory implementation when guidance manuals were published (USEPA/USACE 1991, 1998). USACE initiated research in the early 1990's under the Long-Term Effects of Dredging Operations program to meet these technical and statutory requirements. This research focused on the development of chronic sublethal toxicity tests. Tests developed under this program included a 28-day test with the estuarine amphipod *Leptocheirus plumulosus* measuring effects on survival, growth, and reproduction, and a 28-day test with the marine polychaete *Neanthes arenaceodentata* measuring survival and growth (Emery et al. 1997; Gray et al. 1998; Dillon, Moore,

and Gibson 1993; Dillon, Moore, and Reish 1994; Dillon, Moore, and Gibson 1994; Bridges and Farrar 1997).

A key element in the development of these tests has been the inclusion of technically sound interpretive guidance. While simple acute survival effects are considered relatively straightforward to interpret, more subtle sublethal effects on growth and reproduction require a higher level of understanding to distinguish statistical significance from biological significance. For example, when does a statistically significant reduction in growth or reproduction become biologically important in maintaining a viable population in the environment? Equally important is an understanding of the potential confounding factors (e.g., grain size, ammonia, etc.) that can affect organism responses and interpretation of test results. Consequently, a great deal of effort has been expended in developing tests with consistent and interpretable end points (Ankley et al. 1994; Bridges et al. 1996; Bridges, Farrar, and Duke 1997; Bridges and Farrar 1997; DeWitt, Ditsworth, and Swartz 1988; Gray et al. 1998; Moore and Dillon 1993; Moore and Farrar 1996).

Once implemented within the regulatory testing program, chronic sublethal tests will be used to determine the suitability of dredged material for open-water disposal. Because of the higher costs of alternative disposal options (confined aquatic or upland) and the role these new tests will have in affecting disposal decisions, it is incumbent upon the regulatory authorities (USACE and USEPA) to ensure the consistency and quality of the predictions these tests provide prior to their regulatory implementation. Developing confidence in the quality of the prediction will require experimental validation that the effects measured in the laboratory tests are predictive of measurable effects in the field. The design of a meaningful field validation study requires extensive understanding of the dredged material disposal process, the impacts to be predicted (e.g., impacts to the benthic community related to sediment-associated contaminants), and careful consideration of factors that may impinge on the interpretation of the results (e.g., natural temporal fluctuations in benthic community structure). This technical note provides a review of field validation efforts for sediment toxicity tests conducted to date as well as recommendations for the design and implementation of future field validation efforts.

PREVIOUS FIELD VALIDATION STUDIES: Field validation studies conducted to date have generally used one of two approaches. There are ***experimental/manipulative*** investigations where contaminated material is tested in laboratory-based bioassays and then the same material is placed in the field and subsequent effects on recruitment and benthic community dynamics are tracked over time. There are also ***correlative/observational*** studies where the relationship between laboratory-based sediment toxicity test results and benthic community indices are compared over an existing contamination gradient in the field. In these more typical field validation efforts, samples are collected along a pollution gradient and examined for alterations in benthic community structure and/or function. Sediment toxicity is then evaluated in laboratory-based toxicity tests to determine whether laboratory toxicity correlates with observed differences in the benthic community. A statistically significant correlation is taken as evidence that the laboratory bioassay is predictive of differences in the benthic community and therefore provides a useful predictive tool for such effects in field-exposed communities.

Experimental/Manipulative Studies. While a number of other studies have used experimental/ manipulative approaches for comparing laboratory and field data, most of these were not designed

nor were they intended as field validation efforts. However, the approaches used in these studies are instructive and as such are included in this review. For example, studies by Tagatz and Ivey (1981) and Tagatz et al. (1987) evaluated the effects of pesticide-spiked sediments on benthic recruitment. Although these studies were not implicitly designed as field validation studies, they did use an experimental/manipulative approach to compare laboratory and field colonization rates. These studies examined the effects of the synthetic pyrethroid pesticide fenvalerate on benthic recruitment. In their 1981 study Tagatz and Ivey placed aquaria in the environment to allow "natural" colonization to occur while a second set of aquaria were allowed to colonize under an unfiltered flowing seawater system in the laboratory. The laboratory-developed communities were then dosed continuously via a flow-through aqueous exposure system for 8 weeks at concentrations of 0.01, 0.1, and 1.0 µg fenvalerate/L. For the field portion of the study, aquaria were retrieved from the field after about 8 weeks and then dosed in the lab for 1 week at concentrations of 0.1, 1.0, and 10.0 µg fenvalerate/L. Results indicated that benthic community structure was significantly altered in both laboratory- and field-derived communities at concentrations of 0.1 and 1.0 µg fenvalerate/L. In a 1987 study, Tagatz et al. modified their earlier approach by evaluating recruitment in aquaria containing fenvalerate-spiked sediments. Sediments were spiked at concentrations of 0.1, 1.0, and 10 µg/g. Six groups of four aquaria each were placed in the field. Each group consisted of an unspiked control, 0.1, 1.0, and 10 µg/g spiked sediment. Sediment cores were obtained at 0, 7, 21, 35, 49, and 56 days for subsequent chemical analysis. After 8 weeks all aquaria were removed from the sediments and returned to the laboratory for benthic community analysis. Results indicated that the 10-µg/g treatment showed significant changes in benthic community structure.

A similar study by Kalke, Duke, and Flint (1982) studied the impact of weathered oil on benthic communities colonized in the laboratory and in situ. Aquaria with clean bedded sediment were established under both laboratory and field conditions and allowed to be colonized by planktonic larvae for a period of 8 weeks. Weathered oil was distributed to both laboratory and field replicate compartments; after 4 additional weeks, the experiments were terminated. Results indicated that no significant effects from the oil treatment were observed for the laboratory-colonized communities, whereas total density of macrobenthos, species composition, and numbers of species were significantly reduced in the field-colonized systems. The authors suggested that the differences observed between laboratory and in situ exposures were due to differences in mechanisms of colonization. In the laboratory studies, colonizing organisms were supplied to the aquaria via a water pump whereas in the in situ studies colonizing organisms settled naturally or moved into the compartments from the surrounding environment. As a consequence, in the in situ studies, densities of colonizing organisms were much higher than in the laboratory studies. Low oxidation-reduction potential measurements in both field and lab samples suggested that the oil had reduced the depth of the oxygenated layer by approximately half, which would adversely affect subsurface benthic production and alter other processes such as nutrient regeneration. This reduction in the oxygenated layer manifested effects in the higher density in situ studies but not in the lower density laboratory colonization experiments.

The Field Verification Program (FVP) initiated by USACE and USEPA in the 1980's represents the largest and most relevant (in terms of dredged material testing) field validation effort conducted to date. The FVP followed an earlier and much larger research program conducted by USACE in

the 1970's known as the Dredged Material Research Program (DMRP). The DMRP laid the foundation for many of the tests and procedures used to assess dredged material today, and the FVP was designed as a follow-on program to refine and improve these tools for regulatory implementation. The FVP was one of the first experimental/manipulative type validation studies and was designed to evaluate these proposed tests for the USACE/USEPA effects-based testing program prior to regulatory implementation. A principal objective of the aquatic portion of the FVP was to verify the predictive accuracy of proposed biological tests by measuring the same response in the same selected species in both laboratory bioassays and in the field. Additional objectives included determining the variability and reproducibility inherent in the test method and evaluating the correlation between tissue residues in field-exposed organisms and effects observed in both the laboratory and the field (Gentile et al. 1987).

Although numerous marine species representing four different phyla were evaluated as part of the FVP including annelids (*N. arenaceodentata*, *Nephtys incisa*), molluscs (*Mytilus edulis*, *Yoldia limatula*, and *Mulinea lateralis*), arthropods (*Ampelisca abdita* and *Mysidopsis bahia*), and fish (*Menidia menidia*, *Cyprinodon variegatus*, *Ammodytes americanus*, *Paralichthys dentatus*, and *Pseudopleuronectes americanus*), only biological testing with *N. incisa*, *M. edulis*, *A. abdita*, and *M. bahia* were included in the field validation portion of the study. The remaining species were evaluated for suitability as test organisms and to assess variability and reproducibility of proposed laboratory bioassay test methods.

The dredged material evaluated in the FVP was collected from Black Rock Harbor (BRH), Bridgeport, CT. A historical dredged material disposal site in Central Long Island Sound (CLIS) was selected as the disposal site for the BRH material and the subsequent field monitoring portion of the FVP. The reference sediment used for comparison was collected from the South Reference site of the CLIS. The BRH dredged material contained moderate to high concentrations of polychlorinated biphenyls (PCBs) (6.4 ppm), polycyclic aromatic hydrocarbons (PAHs) (142 ppm), and metals including copper (2,900 ppm), chromium (1,480 ppm), zinc (1,200 ppm), lead (380 ppm), nickel (140 ppm), cadmium (24 ppm), and mercury (1.7 ppm) (Peddicord 1988; Rogerson, Schimmel, and Hoffman 1985).

Stations used to monitor the biological effects in this study were located along the primary axis of current flow (east to west) at the CLIS disposal site. Stations were located at the center of the mound and 200, 400, and 1,000 m to the east along the primary axis of current flow to represent a gradient of decreasing exposure (Figure 1). The site was monitored at these locations for a year preceding disposal of the dredged material and for 3 years post-disposal. Results of acute and chronic laboratory tests conducted on the material just prior to disposal were compared with effects observed in the field following disposal. In addition, material was periodically collected from the disposal mound and evaluated using the laboratory tests to determine whether laboratory test results correlated with changes in field response.

Of the toxicity tests/end points evaluated, measures of mortality, scope for growth (SFG), reproduction, and population response were recommended for predisposal permit testing. Of these end points, only mortality and SFG could be directly validated in field-exposed organisms. Results from laboratory tests measuring survival indicated that only tests with the amphipod *A. abdita* showed significant mortality directly related to the percentage of BRH sediment. Additionally, laboratory

survival was directly related to recolonization of *A. abdita* at the CLIS site indicating that the test was predictive of effects occurring in the field. Similarly, SFG in the mussel *M. edulis* also showed good correspondence between laboratory and field measurements in animals exposed to SPP concentrations of BRH sediment showing a response threshold of about 1.5 mg/L BRH sediment. Furthermore, *M. edulis* SFG showed a strong relationship with accumulated contaminant tissue concentrations. Reproduction and population growth were found to be sensitive end points in both *M. bahia* and *A. abdita*. However, *M. bahia* was not found at the disposal site and *A. abdita* did not recolonize the site in large enough numbers to permit direct field verification of observed laboratory responses for reproduction or population growth.

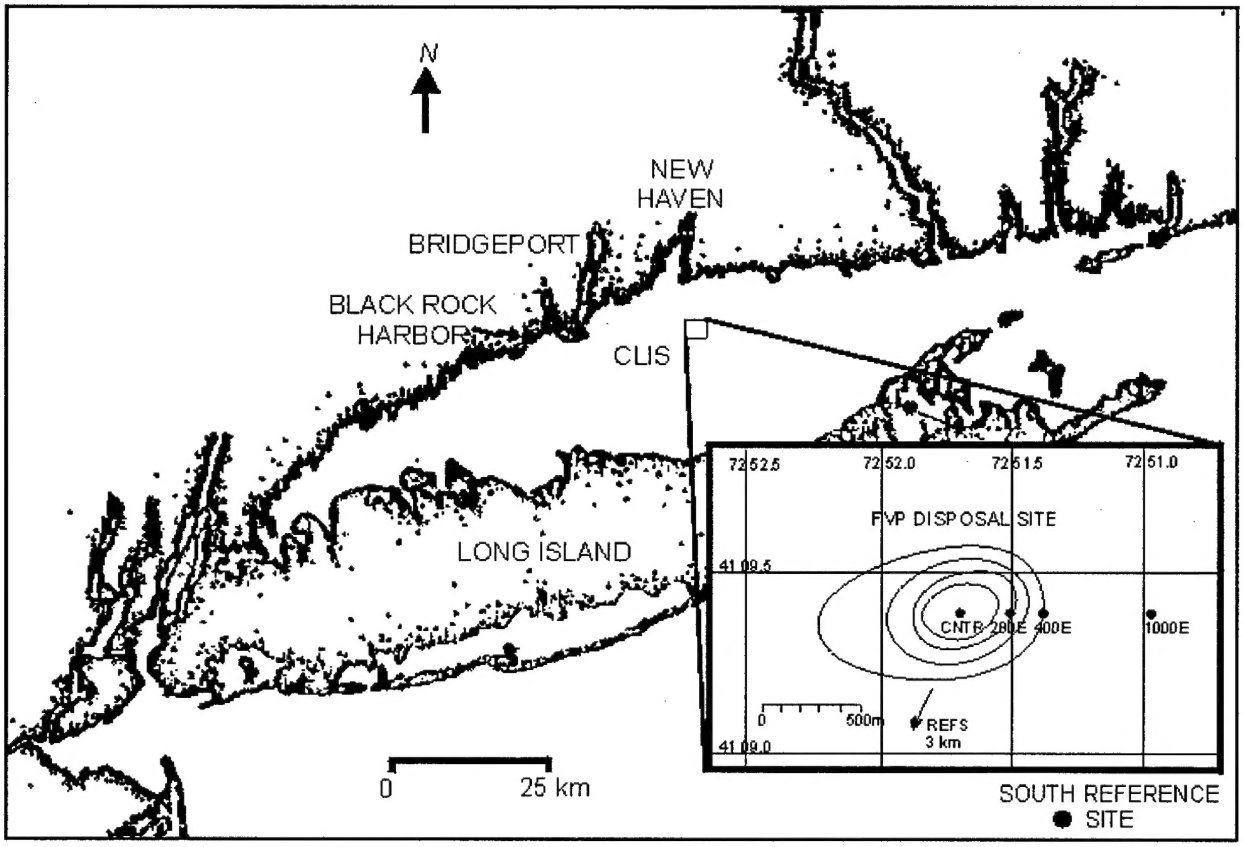

Figure 1. Station locations at CLIS Disposal Site for the FVP (from Gentile et al. 1987)

In an evaluation of bioaccumulation tests, the FVP compared tissue residues in *N. incisa* and *M. edulis* exposed in 28-day laboratory tests to tissue residues in these same species recovered from the field (Gentile et al. 1988). For the field portion of the study, resident *N. incisa* present at the selected station locations (CNTR, 400E, 1000E, and REFS) were recovered from the field while *M. edulis* were collected from a clean control site and deployed from subsurface buoys 1 m above the sediment surface at these same locations and then collected after 28 days of exposure. Measurements of field exposures were based on water column concentrations of PCBs measured at each of the selected field locations. In the laboratory *N. incisa* were exposed for up to 56 days in both bedded sediment and suspended sediment while *M. edulis* were exposed only to suspended sediments in 28-day exposures. The FVP focused exclusively on PCB bioaccumulation to validate the two laboratory-based bioaccumulation tests. PCBs were selected for the field verification component because "of

their persistence, equilibrium partitioning, and kinetics" (Gentile et al. 1988). Comparison of tissue residues from these laboratory-based bioassays with field-collected animals indicated good correlation between laboratory and field exposures for both *N. incisa* (Figure 2a) and *M. edulis* (Figure 2b). The author suggests based on these results that both tests offered reasonable predictions of potential bioaccumulation in the field, at least with regard to PCBs.

Correlative/Observational Studies. A number of studies published by Swartz et al. have used the correlative approach described previously to demonstrate the utility of solid-phase acute amphipod bioassays in predicting the existing conditions of the infaunal community associated with in-place contaminated sediments (Swartz et al. 1982, 1985, 1986, and 1994). In the 1982 study, the authors evaluated sediment toxicity of 175 sediment samples collected from Commencement Bay, Washington, to the amphipod, *Rhepoxynius abronius*. Amphipod survival (10-day) was compared with sediment chemistry and community structure data of in-place sediments. Results indicated a significant correlation between amphipod distribution and sediment toxicity.

In studies published in 1985 and 1986, Swartz et al. evaluated the temporal and spatial extent of impacts from the Los Angeles County Sanitation Districts' sewage outfalls on the benthic community of the Palos Verdes Shelf. Sampling surveys were conducted in the late spring/early summer of 1980 and 1983. A total of eight sampling stations located at increasing distances from 1 to 15 km downcurrent of the outfall diffusers along the 60-m depth contour were sampled. In the 1980 sampling Swartz et al. found that the three stations closest to the outfall (stations 1-3) showed statistically significant acute toxicity relative to the control (98 percent survival). There were no differences in survival among these three stations (average survival equaled 80 percent). In the 1983 sampling survey, survival ranged from 83 to 97 percent for seven of the eight sampling stations (station 2 was not sampled) with no differences relative to control (97 percent). The authors showed that the reduction in acute toxicity to the amphipod *R. abronius* between 1980 and 1983 correlated with a recovery in the macrobenthos and a reduction in the mass emission of biological oxygen demand and chemical contaminants from the outfall. However, Swartz et al. were careful to caution, "The absence of acute sediment toxicity does not demonstrate the absence of benthic degradation." The authors noted that while some amphipods appeared in samples for the first time in the 1983 survey, phoxocephalid amphipods were no closer to the outfall than station 4 (approximately 5 km downcurrent of the outfall diffuser). This finding suggests that effects were occurring at contaminant concentrations lower than would produce effects in the acute tests.

In a similar study published in 1994, Swartz et al. evaluated sediment toxicity to the amphipod *Eohaustorius estuarius*, contaminant concentrations, and amphipod abundance along a contamination gradient in Richmond Harbor, upper San Francisco Bay. The predominant contaminants were dichlorodiphenyltrichloroethane (DDT) and dieldrin, which were produced at a facility in the area from 1946 to 1966. Nine stations showing a gradient in ΣDDT (11.6 to 77,700 g/dry kg) and dieldrin (0.63 to 748 g/dry kg) concentrations were evaluated for amphipod toxicity and amphipod abundance. An evaluation of the ratio of acid volatile solids to simultaneously extracted metals, and concentrations of dieldrin, ΣPAHs, and Arochlor 1254 suggested that concentrations of these constituents were too low to affect toxicity in *E. estuarius* and that the concentrations of ΣDDT were sufficient to result in acute toxicity. Further, the authors found a correlation ($R^2 = 0.72$, $p = 0.004$ for the Lauritzen Channel/Richmond Harbor sediments) between increasing carbon-normalized

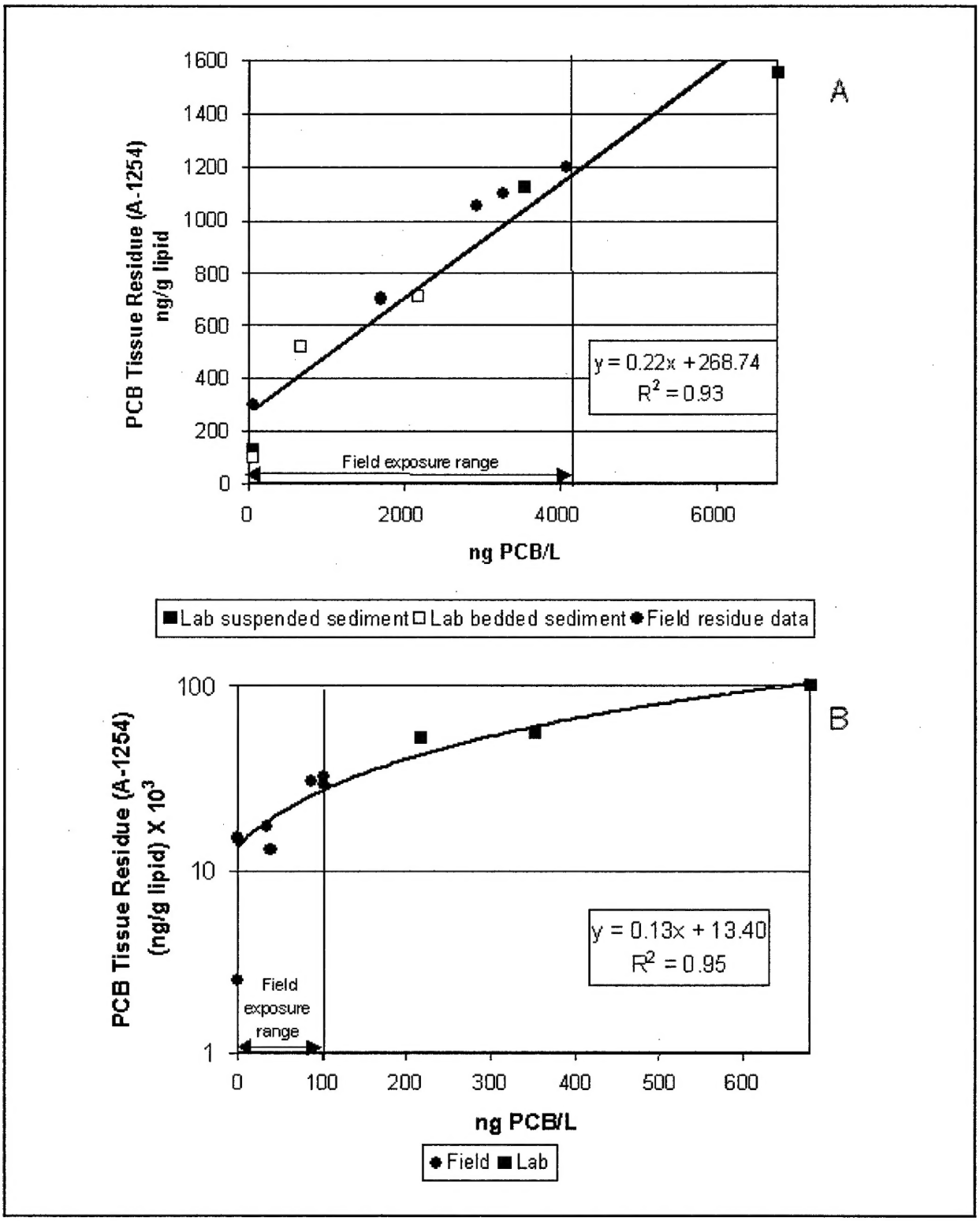

Figure 2. PCB tissue residues (as Aroclor 1254) in both laboratory and field organisms exposed to BRH sediment; A) *Nephtys incisa* exposed to bedded sediment; B) *Mytilus edulis* exposed to suspended sediment only (field exposures for *M. edulis* occurred at 1 m above bottom) (Field exposure range based on measured concentrations of PCBs at the selected station locations) (adapted from Gentile et al. 1988).

ΣDDT concentration and mortality of the amphipod *E. estuarius* in 10-day laboratory sediment toxicity tests (Figure 3). An analysis of amphipod abundance along the same contamination gradient failed to show a statistically significant correlation (R^2 = 0.19, p = 0.23 for the Lauritzen Channel/Richmond Harbor sediments) (Figure 4). One amphipod species, *Grandidiarella japonica*, showed increasing abundance with increasing ΣDDT concentration. It was not clear why the abundance of *G. japonica* was positively correlated with sediment toxicity and ΣDDT concentrations since this species shows similar sensitivity in aqueous exposures to reference toxicants compared with other amphipod species. It is possible, although not suggested by the authors, that the benthic communities present at the sites may have adapted to contaminant levels in these sediments. If so, the potential for adaptation may present a challenge to the correlative/observational approach.

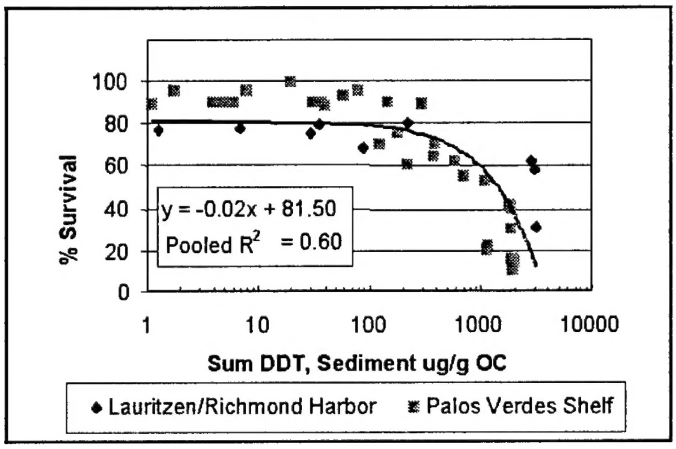

Figure 3. Mean survival of *E. estuarius* in Lauritzen Channel/Richmond Harbor sediment and *R. abronius* in Palos Verdes Shelf sediment in relation to the carbon-normalized ΣDDT concentration in sediment (OC = organic carbon) (adapted from Swartz et al. 1994)

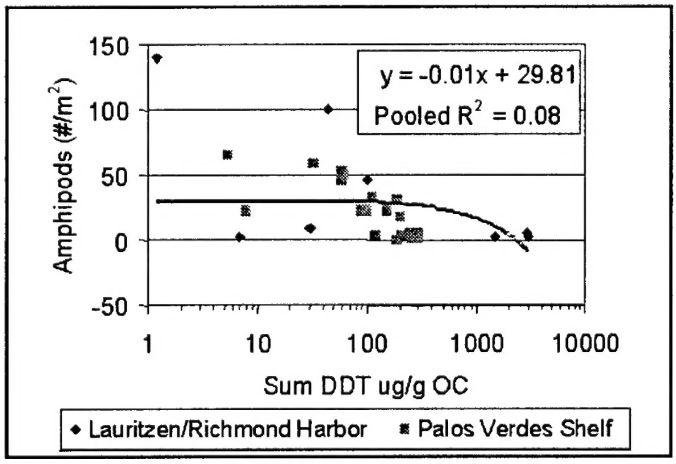

Figure 4. Mean abundance of amphipods (excluding *G. japonica*) in relation to ΣDDT normalized to organic carbon for Lauritzen/Richmond Harbor and Palos Verdes Shelf sediments (adapted from Swartz et al. 1994)

Finally, pooling data with their earlier studies of a contamination gradient on the Palos Verdes Shelf, Swartz et al. (1994) showed similar relationships for sediment toxicity (Pooled R^2 = 0.60, p << 0.001) and amphipod abundance (Pooled R^2 = 0.08, p = 0.11) compared with organic carbon normalized ΣDDT sediment concentrations (Figures 3 and 4) (Swartz et al. 1994).

McGee and Fisher (1999) and McGee et al. (1999) also used the correlative approach to evaluate the relationship between sediment-associated contaminants, abundance of the amphipod *L. plumulosus*, and toxicity to *L. plumulosus* in 10-day laboratory-based bioassays. The study evaluated

sediments from 25 locations in Baltimore Harbor. Sediments had effects range median quotients (ERM-Qs)[1] ranging from <1 to >20. Principal contaminants included metals, PAHs, and PCBs. Survival in acute toxicity tests with these sediments ranged from 0 to >129 percent with seven of the sites showing statistically reduced survival relative to control. For six of the test samples where survival was in excess of 100 percent the authors indicated that the excess was likely due to the presence of indigenous amphipods in the samples. It is interesting to note that individual replicates for the toxicity tests were actually derived from separate independent grabs at a given location. This is a departure from the standard approach used in sediment toxicity tests where a single homogenized sample is divided among the laboratory replicates for the test. Only one of these independent grab samples was evaluated for chemistry.

A plot of the 10-day toxicity test results (including the retest data) against field densities of amphipods at the test sites suggests a positive correlation ($R^2 = 0.41$, $p = 0.001$) between amphipod density and amphipod survival (Figure 5a). The correlation improved slightly ($R^2 = 0.52$, $p = 0.0002$) with exclusion of a potential outlier value (Figure 5b). In addition, the authors showed a weak negative correlation ($R^2 = 0.19$, $p = 0.03$) with concentrations of contaminants (expressed as ΣERM-Q)[2] and amphipod survival (Figure 6a). Again the correlation improved slightly ($R^2 = 0.29$, $p = 0.006$) with exclusion of a potential outlier (Figure 6b). It is important to note that correlations might have been stronger if spatial variability in the field had not been incorporated into the sampling design via the use of independent grab samples for each laboratory toxicity replicate and had chemical analysis been performed on a composite or all of the sediments rather than a single replicate.

In a separate study, McGee and Fisher (1999) summarized results of chronic sublethal tests with *L. plumulosus* that were conducted in parallel with acute tests on selected samples from many of the same sites evaluated in their earlier study. Unfortunately, nearly all of the 11 sites included in this study were depauperate of *L. plumulosus*. Consequently, a comparison between laboratory response and field densities of *L. plumulosus* was not meaningful. A comparison of survival in these laboratory tests to broader indices of benthic community health (e.g., number of taxa, total number of organisms, Shannon-Weaver diversity index, and the index of biological integrity (IBI)) also showed little to no relationship (Figure 7). Similarly, a comparison of sublethal end points (e.g., growth and reproduction) to these same indices also showed no relationship. It is not clear why this was the case. Perhaps the contamination gradient was not broad enough, or factors other than contaminants were affecting differences in the benthic communities. Alternatively, it could be that the community of organisms, showing no gross changes in total abundance, species richness, or other indices, was populated by organisms that were less sensitive or less exposed to the sediment contaminants.

---

1 The effects range median quotient (ERM-Q) is derived by dividing the actual measured concentration in the sediment by the corresponding effects range median (ERM) value for that particular contaminant. ERM values were developed by Long et al. (1995) using the distribution of observed mortality in laboratory bioassays for a large number of sediments collected from a variety of areas. The ERM value represents the concentration of a particular contaminant that causes significant mortality in the test species 50 percent of the time for the range of sediments used in deriving the value.

2 ΣERM-Q represents the sum of all ERM-Q values for a particular sediment.

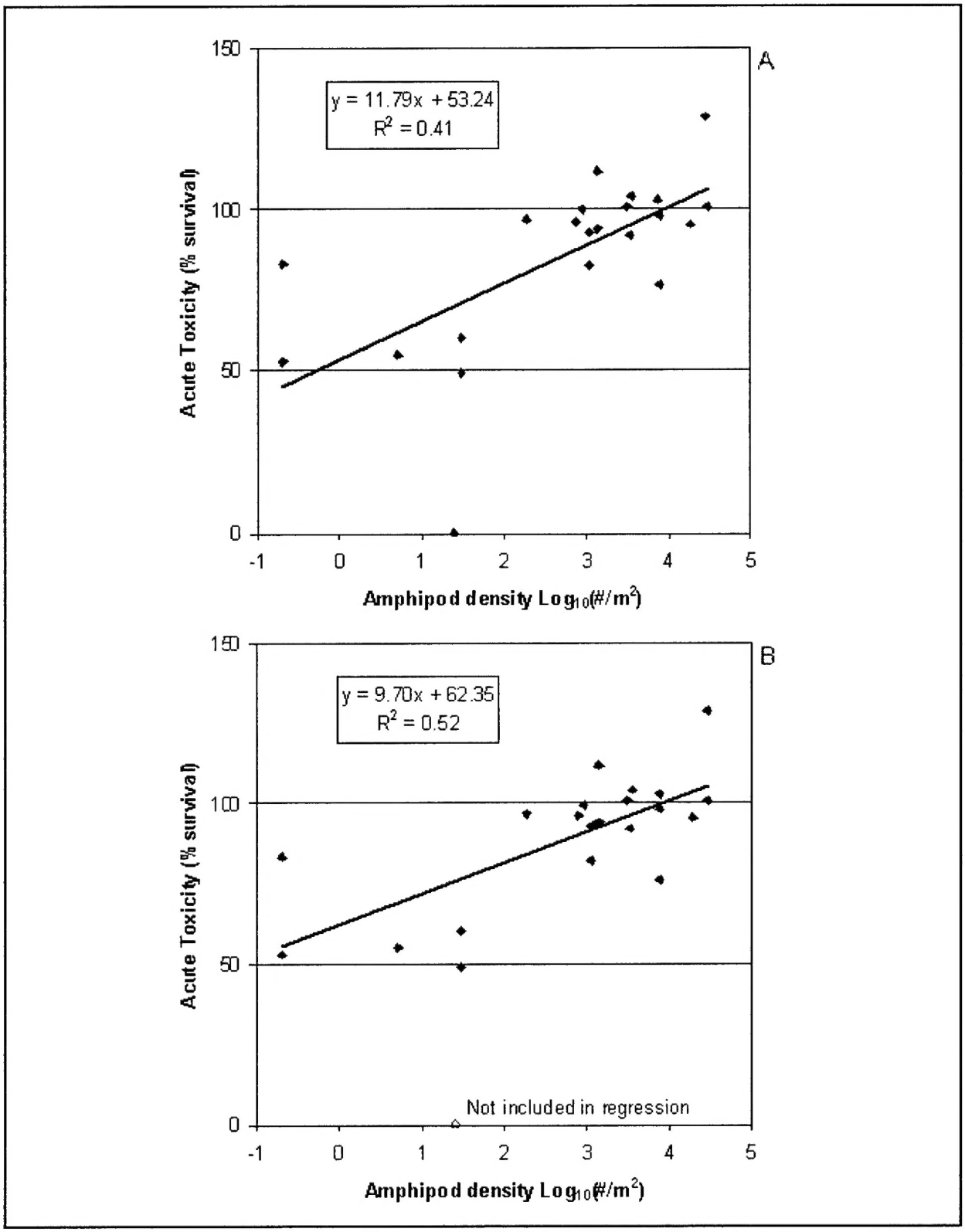

Figure 5. Relationship between survival in 10-day acute toxicity tests with *L. plumulosus* and field densities of *L. plumulosus* at selected sites in Baltimore Harbor A) with a potential outlier, and B) without a potential outlier (based on data from McGee et al. 1999)

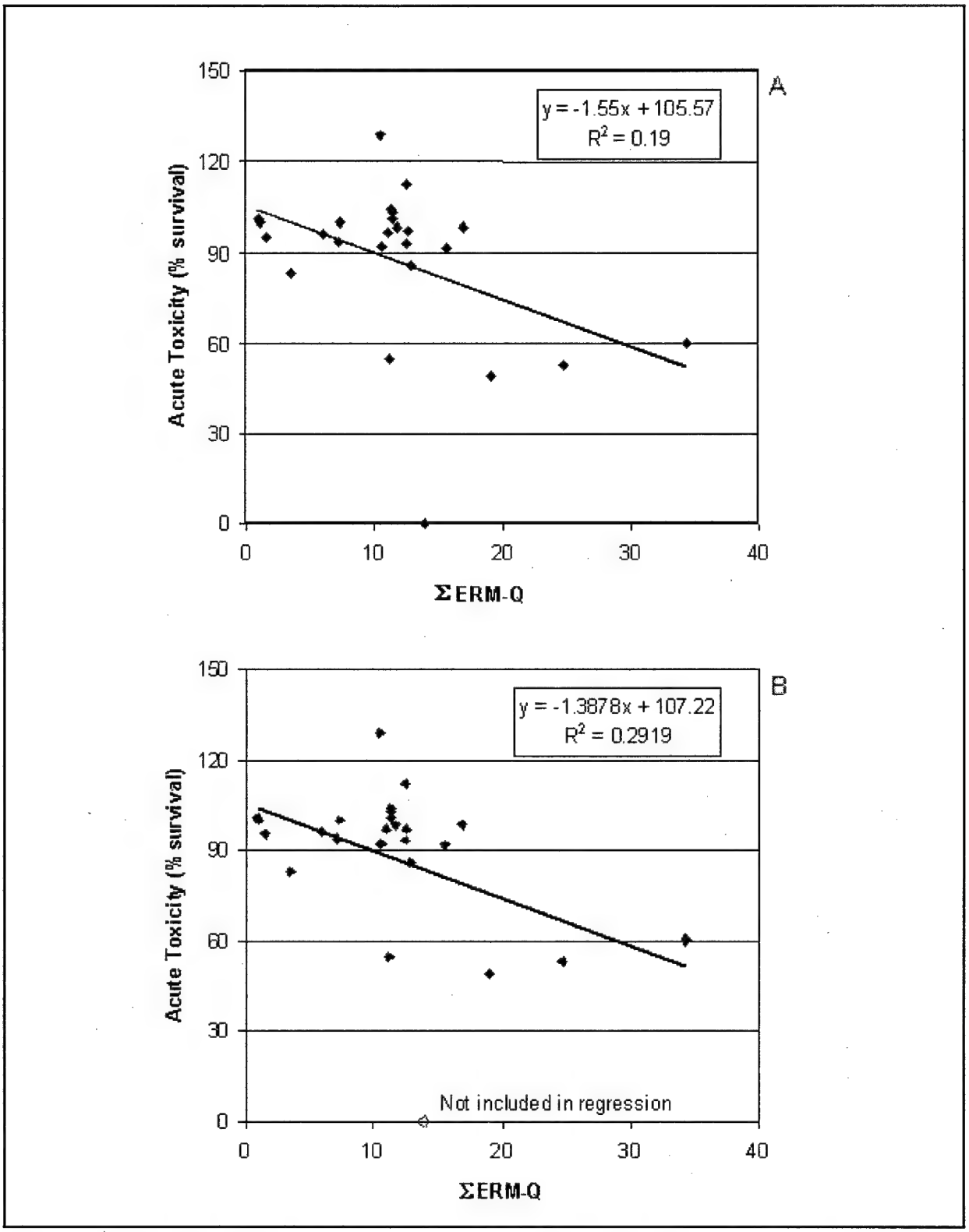

Figure 6. Relationship between survival in 10-day acute toxicity tests with *L. plumulosus* and ΣERM-Q A) with a potential outlier, and B) without a potential outlier (based on data from McGee et al. 1999)

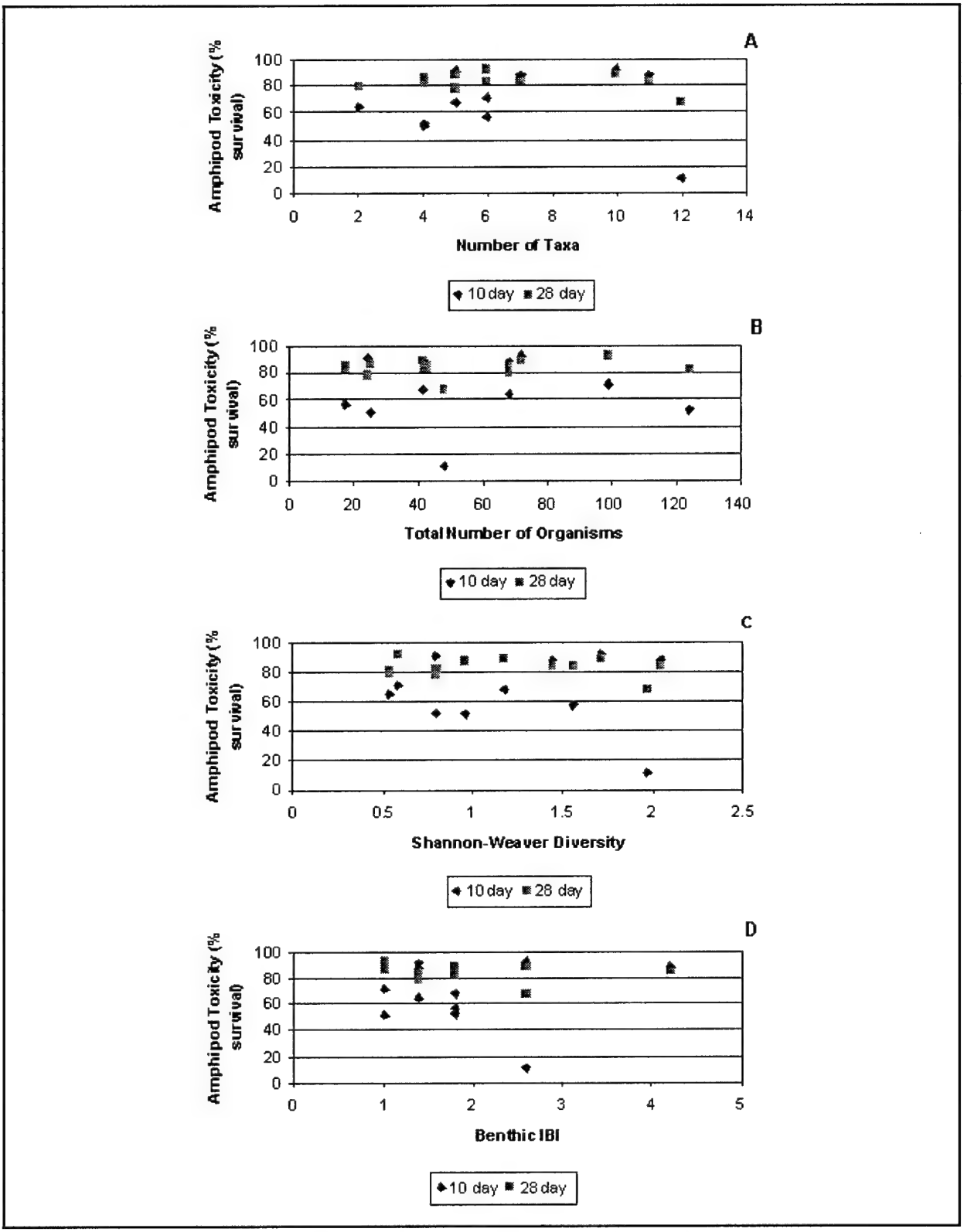

Figure 7. Relationship between 10- and 28-day survival and benthic community indices: A) number of taxa; B) total number of organisms; C) Shannon-Weaver diversity index, and D) benthic IBI (based on data from McGee and Fisher 1999)

Brunson et al. (1998) conducted a correlative type study designed, in part, to validate the use of the oligochaete worm *Lumbriculus variegatus* as a bioaccumulation test species for freshwater sediments. Sediment samples and resident oligochaetes were collected from 13 navigational pools in the Upper Mississippi and the St. Croix Rivers. Tissue residues were measured in *L. variegatus* after 28 days of exposure to field-collected sediment and compared with tissue residues in field-collected animals. While results of tissue analysis indicated that both PAHs and PCBs were consistently found at concentrations above detection limits, the authors focused on the PAH data in their evaluation of the 28-day *L. variegatus* bioaccumulation test. Lipid-normalized tissue concentrations of PAHs showed a positive correlation between laboratory-exposed and field-collected organisms. Generally there was a change in tissue residue concentrations for individual PAHs; field tissue residues were greater than laboratory tissue residues for the lower molecular weight PAH tissue residues, and laboratory tissue residues were greater than field tissue residues for the high molecular weight PAHs. The authors speculated that the observed differences between laboratory and field-exposed animals may have been a result of any one of a number of factors. For example, it is possible that the more water soluble lower molecular weight PAHs were lost during sampling of sediments for the laboratory exposures. Differences may also have arisen from the spatial heterogeneity of contaminants in the field (e.g., grab samples may have included sediment from depths to which field-collected organisms were not exposed). Another possibility may be related to the fact that field organisms may be exposed through a number of routes (e.g., sediment, water, and food) whereas the primary route of exposure in laboratory bioaccumulation tests is through direct contact and ingestion of sediment. There may also have been genetic/phenotypic or possibly species differences between the lab and field organisms examined that may have lead to differences in uptake. However, despite these differences, the authors indicate that the ratio of tissue concentrations for specific PAHs in laboratory and field-exposed organisms were generally similar with about 90 percent of the corresponding concentrations for laboratory versus field-collected organisms being within a factor of 2-3. In addition biota to sediment accumulation factors (BSAFs) derived for lab and field data showed good correlation with the field-collected organisms, having BSAF values equal to or higher than the laboratory ($R^2 = 0.80$, $p = 0.003$). The good correlation indicates that laboratory bioaccumulation tests with the oligochaete *L. variegatus* were good predictors of bioavailability (Figure 8). Based on these findings the authors conclude, "Laboratory results could be extrapolated to the field with a reasonable degree of certainty."

CONCLUSIONS AND RECOMMENDATIONS: Each of the two basic approaches employed to date to field-validate laboratory bioassays (i.e., experimental/manipulative versus correlative/observational) has inherent advantages and disadvantages. In general, correlative approaches are subject to greater variability in the field portion of the assessment. In correlative studies, the spatial variability of contaminants and contaminant mixtures remains intact. As one moves along the "contaminant gradient," the type and amount of co-occurring contaminants change as do other physical/chemical features of the sediment matrix and the biological communities that respond to these factors. The difficulty in trying to ascribe differences in benthic community structure to contaminant levels is evidenced by the low correlations found by Swartz et al. (1994) (Figure 4) and McGee and Fisher (1999) (Figure 7). While spatial variability is inherent in environmental contamination, such variability can confound attempts to validate responses measured in laboratory tests where such variability is attenuated through homogenization of the sample matrix. The experimental/manipulative approach attempts to overcome this variability via sample compositing and/or homogenization to remove/reduce spatial variation.

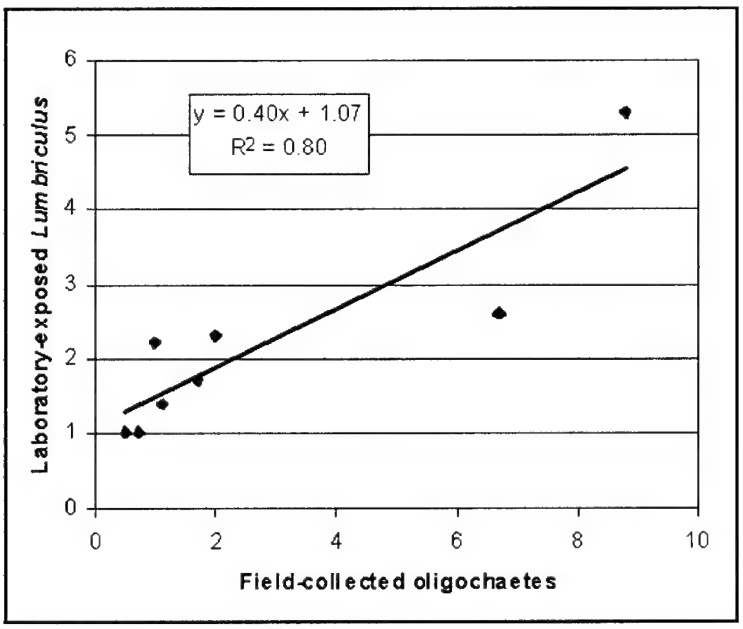

Figure 8. Biota to sediment accumulation factors (BSAFs) for PAHs in laboratory-exposed *L. variegatus* compared with BSAFs for field-collected oligochaetes (adapted from Brunson et al. 1998)

In the FVP, sediment was placed in the field in a way that reflected how the material would be handled in a dredged material assessment. It is unlikely in the current regulatory climate that a similar study could be conducted today (i.e., contaminated material placed in the environment in an unconfined manner without subsequent capping). Consequently, future experimental/manipulative studies will most likely rely on the use of sediment placed in small, recoverable containers in the field. The use of containers can lead to effects that may not necessarily reflect effects in unenclosed systems. It is possible that using containers to evaluate contaminated sediment effects on benthic recolonization may lead to effects on recruitment that do not necessarily reflect recruitment patterns of the surrounding community. Dudzik et al. (1979) reported heightened biological and chemical activity along the edges and bottoms of experimental microcosms. Stephenson et al. (1984) showed differences in the distribution of planktonic organisms related to the size of enclosure. Flemmer et al. (1993) demonstrated effects of container size on benthic recolonization in laboratory microcosms. These "edge effects" where spurious recruitment patterns are observed at the edge of the container and issues associated with scaling (i.e., container size) must be considered and accounted for in these types of studies.

Choosing between an experimental/manipulative and a correlative/observational approach or development of new approaches for future validation efforts must be driven by the questions being asked. For example, if one is interested in evaluating the predictive ability of a test for the assessment of in-place contaminants, a correlative approach might be more appropriate as this more closely approximates the type of assessment one is trying to represent. However, if one is trying to validate a test for assessing potential effects associated with a disposal activity where material is physically displaced to a new location (e.g., dredging and disposal), then an experimental approach might more closely approximate the scenario one is trying to evaluate.

In both correlative/observational and experimental/manipulative approaches it is important to measure a variety of biological (abundance, diversity, size/age class) and physicochemical parameters (sediment chemistry, grain size, total organic carbon, redox, etc.) to capture important changes in biological response and potential factors affecting exposure. Consideration must be given to understanding the dynamics of the community (e.g., patterns of recruitment) in the area where the study will be conducted. This will help to ensure that the field assessment is conducted in a way that captures changes in the benthic community and enhances the ability of the study to discriminate meaningful differences in community level responses. Because changes in benthic community

structure and function may not be manifest in a single season or even a single year, it is important for experimental/manipulative studies to monitor effects in the field with a frequency and duration that are sufficient to capture these changes. This may require monitoring sites more than once per year and a study duration in excess of 2 years (Underwood 1993).

Because of the spatial heterogeneity of contaminants in correlative studies and the inherent variability associated with measuring responses in the field, it is important to have a replicated design. A minimum of two, preferably three, replicate measurements should be made to ensure that statistically meaningful differences are detected (Sanders 1984; Underwood 1993).

Since most field validation efforts are designed to evaluate contaminant-induced effects, it is also important to account for factors that may influence the availability of the contaminants. Establishing a correlation between measured contaminant levels and responses in either field- or laboratory-exposed organisms can be facilitated by normalizing for factors that control bioavailability. Most field validation studies performed to date have normalized concentrations of organic contaminants to organic carbon levels present in the sediment. Similarly, selected divalent metals can be normalized to the concentration of acid volatile sulfides. A separate but related issue, especially for correlative studies, is accounting for differences in the magnitude and ratio of individual contaminants present in the typical milieu of contaminants found in sediments across a range of sites. Typically, quotient approaches are used in an attempt to account for these differences. Quotient approaches assume additivity and have perhaps the greatest utility when they are used to evaluate concentrations within a contaminant class (e.g., PAHs). The quotient approach accounts for differences in the potency of the individual constituents within a given contaminant class that has the same mode of action. For example, dioxins (i.e., tetrachloro-dibenzo-*p*-dioxins (TCDD) and dibenzofurans) can be normalized to 2,3,7,8 TCDD toxic equivalents to account for changes in the type and amount of the 14 individual dibenzodioxins and dibenzofuran congeners over a range of sites. Similarly the ΣPAH model of Swartz et al. (1995) can be used to normalize for the types and amount of PAHs over a range of sites. When multiple contaminant classes are present, the ERM-Q approach of Long et al. (1995) can be used to provide qualitative information on the level of contamination and associated toxicity over a range of sites.

The data of Swartz et al. (1994) and McGee and Fisher (1999) showed no significant differences in variety of benthic community measurements along a contamination gradient that produced significant toxicity in laboratory bioassays. These differences may be accounted for in the type of organisms colonizing the sites as they have been either less sensitive or less exposed (i.e., tube-dwellers versus free-burrowing organisms) than the organisms used in the laboratory bioassay. Another possibility is that resident organisms may have adapted to the contaminant loads over time. In either case, field studies and correlative studies in particular must carefully consider the potential masking of contaminant effects by factors such as these. As Swartz et al. (1985) point out in their study of the Palos Verdes outfall, the absence of a specific taxa (e.g., phoxocephalid amphipods) may be as important as more standard measures of benthic community analysis. A comparison of resident and laboratory test organisms of the same species via toxicity testing (reference toxicant or whole sediment) could shed light on whether an adaptive community has evolved in the field.

In correlative/observational studies, it is important to document and account for the distribution of physicochemical features (e.g., current flow and direction, water depth) of the sampling sites along

the gradient and the potential of these features to affect recruitment patterns and potentially confound interpretation of benthic infaunal data. A principal assumption of the correlative approach is that the primary factor affecting differences in benthic community structure and/or function along the gradient is the level of chemical contamination and that all other factors have minimal to no influence on the benthic community. One way to account for differences associated with physical location might be a hybrid of the correlative and experimental/manipulative approaches where representative sediments from each of the sites along the gradient are placed in containers at selected locations along the gradient. In this way, factors associated with a given location would presumably act equally upon all sediment types placed at the location, and effects related to contaminant loading could be discriminated.

Another important consideration is selecting a field site and measuring end points that are relevant to the lab bioassay being validated. The ideal would be to have the test organism be an important if not critical component of the benthic community at the selected field locations. While this may not be possible or even necessary, it is important that the end point/test organism being evaluated at least have relevance to the individual field location. For example, one probably would not want to validate a laboratory test with a test animal that is typically found in fine-grain material at a site where coarse material predominates.

Finally, the potential for physical alterations of the site via wave action during such events as storms or boat traffic should be considered as part of the site selection process in both correlative and experimental studies. The purpose of the field validation effort is to establish whether or not contaminant-induced responses in the laboratory are reflective of contaminant-induced effects in the field. Having a major storm mobilize or bury all the contaminated sediments at a test location obviously affects the ability of the study to validate the laboratory test.

In the field, a variety of physical and chemical processes act upon both sediment-associated contaminants and organisms to constantly alter and affect exposure in a very dynamic manner. However, in the laboratory the sample is intentionally kept under very controlled conditions to minimize changes in these physical and chemical processes on exposure. In addition, animals in the field are subject to competition, predation, and a host of other "naturally occurring stressors" while these stressors are minimized or eliminated in laboratory-based assays. It has also been argued by Swartz (Swartz et al. 1986) and others that acute laboratory-based toxicity tests with single organisms cannot account for degraded benthic communities at lower levels of contamination because changes in benthic community structure at these lower levels of contamination are driven by longer term, sublethal effects. Chronic effects on growth or reproduction may result in shifts in benthic community structure and function. Consequently, it is possible that chronic tests may better predict potential changes in field-exposed communities.

Chronic sublethal measures of sediment toxicity are required under the existing regulatory frame-work for the evaluation of dredged material. Because these tests are more costly and complex to conduct, it is important to ensure that the responses in these chronic sublethal tests correlate with effects measured in the field prior to their regulatory implementation. The application of a properly designed field validation effort offers the best opportunity for establishing confidence in these laboratory-based bioassays.

ACKNOWLEDGMENTS: The author, Dr. David W. Moore, MEC Analytical Systems, Carlsbad, CA, gratefully acknowledges the contributions of Drs. Todd Bridges, Jeffery A. Steevens, and Laura S. Inouye, Environmental Risk Assessment Branch, Environmental Laboratory (EL), U.S. Army Engineer Research and Development Center; Jack Word and Doug Diener, MEC; and Michael R. Palermo, Environmental Processes and Engineering Division, EL; and Mrs. Lucinda Word, MEC, in providing thoughtful comment and review of the manuscript. In addition, Mrs. Word, Mrs. Michelle Patzius, MEC, and Ms. Suzie Watts, MEC, provided invaluable assistance in preparing the manuscript for publication.

POINT OF CONTACT: For additional information, contact Dr. Todd S. Bridges (601-634-3626, *Todd.S.Bridges@erdc.usace.army.mil*) or the Program Manager of the Dredging Operations and Environment Research Program, Dr. Robert M. Engler (601-634-3624, *Robert.M.Engler@erdc.usace.army.mil*). This technical note should be cited as follows:

> Moore, D. W. (2001). "Review of field validation studies of sediment bioassays for the regulatory evaluation of dredged material," *DOER Technical Notes Collection* (ERDC TN-DOER-C23), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/dots/doer/*

REFERENCES

Ankley, G. T., Benoit, D. A., Balogh, J. C., Reynoldson, T. B., Day, K. E., and Hoke, R. A. (1994). "Evaluation of potential confounding factors in sediment toxicity tests with three freshwater invertebrates," *Environ. Toxicol. Chem.* 13, 627-635.

Bridges, T. S., and Farrar, J. D. (1997). "The influence of worm age, duration of exposure, and endpoint selection on bioassay sensitivity for *Neanthes arenaceodentata* (Annelida: Polychaeta)," *Environ. Toxicol. Chem.* 16(8), 1650-1658.

Bridges, T. S., Farrar, J. D., and Duke, B. M. (1997) "The influence of food ration on sediment toxicity in *Neanthes arenaceodentata* (Annelida: polychaeta)," *Environ. Toxicol. Chem.*16(8), 1659-1665.

Bridges, T. S., Farrar, J. D., Gamble, E. V., and Dillon, T. M. (1996). "Intraspecific density effects in *Nereis (Neanthes) arenaceodentata* Moore (Polychaeta: Nereidae)," *Jnl of Exp Mar Biol & Ecology* 19(5), 221-235.

Brunson, E. L., Canfield, T. J., Dwyer, F. J., Ingersoll C. G., Kemble, N. E. (1998). "Assessing the bioaccumulation of contaminants from sediments of the upper Mississippi River using field-collected oligochaetes and laboratory-exposed *Lumbriculus variegatus*," *Arch. Environ. Contam. Toxicol.* 35, 191-201.

DeWitt, T. H., Ditsworth, G. R., and Swartz, R. C. (1988). "Effects of natural sediment features on survival of the phoxocephalid amphipod *Rhepoxynius abronius*," *Mar. Env. Res.* 25, 99-124.

Dillon, T. M., Moore, D. W., and Gibson, A. B. (1993). "Development of a chronic sublethal bioassay for evaluating contaminated sediment with the marine polychaete worm, *Nereis (Neanthes) arenaceodentata*," *Environ. Toxicol. Chem.* 12, 589-605.

Dillon, T. M., Moore, D. W., and Gibson, A. B. (1994). "Development of a chronic sublethal bioassay for evaluating contaminated sediment with the marine polychaete worm *Nereis (Neanthes) arenaceodentata*," Miscellaneous Paper D-94-5, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Dillon, T. M., Moore, D. W., and Reish, D. J. (1994). "A 28 day sediment bioassay with the marine polychaete, *Nereis (Neanthes) arenaceodentata*," *Environmental toxicology and risk assessment* - 3rd Volume, ASTM STP 1218, J. S. Hughes, G. R. Biddinger, and E. Mones, ed., American Society for Testing Materials, West Conshohocken, PA.

Dudzik, M., Harte, J., Jassby, A., Lapan, E., Levy, D., and Rees, J. (1979). "Some considerations in the design of aquatic microcosms for plankton studies," *Int. J. Environ. Stud.* 2(13), 125-130.

Emery, V. L., Jr., Moore, D. W., Gibson, A. B., Gray, B. R., Duke, B. M., Wright, R. B., and Farrar, J. D. (1997). "Development of a chronic sublethal sediment bioassay using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker)," *Environ. Toxicol. Chem.* 16(9), 1912-1920.

Flemmer, D. A., Clark, J. R., Stanley, R. S., Bundrick, C. M., and Plaia, G. R. (1993). "Importance of physical scaling factors to benthic marine invertebrate recolonization of laboratory microcosms," EPA/600/J-94/114, *Int. J. Environ. Stud.* 44, 161-179.

Gentile, J. H., Scott, K. J., Lussier, S. M., and Redmond, M. S. (1987). "The assessment of Black Rock Harbor dredged material impacts on laboratory population responses," Technical Report D-87-3, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Gentile, J. H., Pesch, G. G., Lake, J., Yevich, P. P., Zaroogian, G., Rogerson, P., Paul, J., Galloway, W., Scott, K., Nelson, W., Johns, D., and Munns, W. (1988). "Synthesis of research results: Applicability and field verification of predictive methodologies for aquatic dredged material disposal," Technical Report D-88-5, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Gray, B. R., Emery, V. L., Jr., Brandon, D. L., Wright, R. B., Duke, B. M., Wright, R. B., and Farrar J. D, and Moore, D. W. (1998). "Selection of optimal measures of growth and reproduction for the sublethal *Leptocheirus plumulosus* sediment bioassay," *Environ. Toxicol. Chem.* 17(11), 2288-2297.

Kalke, R. D., Duke, T. W., and Flint, R. W. (1982). "Weathered IXTOC-I oil effects on estuarine benthos," *Estuarine, Coastal, and Shelf Science* 15, 75-84.

Long, E. R., MacDonald, D. D., Smith, S. L., and Calder, F. D. (1995). "Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments," *Environmental Management* 19, 81-97

McGee, B. L., and Fisher, D. J. (1999). "Final technical report: Field validation of the chronic sediment bioassay with the estuarine amphipod *Leptocheirus plumulosus* in Chesapeake Bay," prepared for U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

McGee, B. L., Fisher, D. J., Yonkos, L. T., Ziegler, G. P., and Turley, S. (1999). "Assessment of sediment contamination, acute toxicity, and population viability of the estuarine amphipod *Leptocheirus plumulosus* in Baltimore Harbor, Maryland," *Environ. Toxicol. Chem.* 18(10), 2151-2160.

Moore, D. W., and Dillon, T. M. (1993). "The relationship between growth and reproduction in the marine polychaete worm, *Neanthes arenaceodentata*: Implications for chronic sublethal sediment bioassays," *Jnl Exp Mar Biol & Ecol* 173, 231-246.

Moore, D. W., and Farrar, J. D. (1996). "Effect of growth on reproduction in the freshwater amphipod *Hyalella azteca* (Saussure)," *Hydrobiologia* 328, 127-134.

Peddicord, R. K. (1988). "Summary of the US Army Corps of Engineers/US Environmental Protection Agency Field Verification Program," Technical Report D-88-6, prepared by Environmental Research Laboratory, U.S. Environmental Protection Agency, Narragansett, RI, for the U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Rogerson, P. F., Schimmel, S. C., and Hoffman, G. (1985). "Chemical and biological characterization of Black Rock Harbor dredged material," Technical Report D-85-9, prepared by Environmental Research Laboratory, U.S. Environmental Protection Agency, Narragansett, RI, for the U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS. NTIS No. AD A164 065.

Sanders, W. M. (1984). "Field validation," *Fundamentals of aquatic toxicology*. G. M. Rand and S. R. Petrocelli, ed., Hemisphere Publishing, Washington, Chapter 21, 601-618.

Stephenson, G. L., Hamilton, P., Kaushik, N. K., Robinson, J. B., and Solomon, K. R. (1984). "Spatial distribution of plankton in enclosures of three sizes," *Can. J. Fish. Aquat. Sci.* 7(41), 1048-1054.

Swartz, R. C., Cole, F. A., Lamberson, J. O., Ferraro, S. P., Schults, D. W., Waldemar, H. L., II, DeBen, W. A., and Ozretich, R. J. (1994). "Sediment toxicity, contamination and amphipod abundance at a DDT- and dieldrin-contaminated site in San Francisco Bay," *Environ. Toxicol. Chem.* 13, 949-962.

Swartz, R. C., Cole, F. A., Schults, D. W., and DeBen, W. A. (1986). "Ecological changes in the Southern California Bight near a large sewage outfall: Benthic conditions in 1980 and 1983," *Marine Ecology Progress Series* 31, 1-13.

Swartz, R. C., DeBen, W. A., Sercu, K. A., and Lamberson, J. O. (1982). "Sediment toxicity and the distribution of amphipods in Commencement Bay, Washington, USA," *Mar. Pollut. Bull.* 13(10), 359-364.

Swartz, R. C., Schultz, D. W., Ditsworth, G. R., DeBen, W. A., and Cole, F. A. (1985). "Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall," *Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems*. ASTM STP 865, T. P. Boyle, ed., American Society for Testing and Materials, West Conshohocken, PA, 152-175.

Swartz, R. C., Schults, D. W., Ozretich, R. J., Laberson, J. O., Cole, F. A., DeWitt, T. H., Redmond, M. S., and Ferraro, S. P. (1995). "SPAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected sediments," *Environmental Toxicology and Chemistry* 141977-1987.

Tagatz, M. E., and Ivey, J. M. (1981). "Effects of fenvalerate on field and laboratory-developed estuarine benthic communities," *Bull. Environ. Contam. Toxicol.* 27, 256-267.

Tagatz, M. E., Stanley, R. S., Plaia, G. R., and Deans, C. H. (1987). "Responses of estuarine macrofauna colonizing sediments contaminated with fenvalerate," *Environ. Toxicol. Chem.* 6, 21-25.

Underwood, A. J. (1993). "The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world," *Australian Journal of Ecology* 18, 99-116.

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers. (1991). "Evaluation of dredged material proposed for ocean disposal - Testing manual," EPA-503/8-91/001, Washington, DC.

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers. (1998). "Evaluation of dredged material proposed for discharge in waters of the U.S.- Testing manual," EPA-823-B-98-004, Washington, DC.